A Analysis

A.1 Preliminaries

We present two well-known inequalities that connect the concept of expected supremum to the metric entropy of T denoted by $\log \mathcal{N}(T, d, \epsilon)$.

First, we state Dudley's inequality, which as stated by [49] gives the supremum of X_t in terms of the metric entropy of T denoted by $\log \mathcal{N}(T, d, \epsilon)$.

Theorem 6 (Dudley's integral inequality)**.** *Let $(X_t)_{t\in T}$ be a mean-zero random process on a metric space (T, d) with sub-Gaussian increments. Then, there exists a constant $C > 0$ such that*

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \leq CK \int_0^\infty \sqrt{\log \mathcal{N}(T, d, \epsilon)}d\epsilon.$$

We refer the reader to [49] for a detailed discussion and a proof of the theorem.

One can also obtain a lower bound on $\mathbb{E}\left[\sup_{t\in T} X_t\right]$ by Sudakov's inequality (see Theorem 7.4.1 by [49]) stated below.

Theorem 7 (Sudakov's inequality)**.** *Let $(X_t)_{t\in T}$ be a mean-zero Gaussian process on a metric space (T, d). Then, for any $\epsilon > 0$, we have*

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \geq c\epsilon\sqrt{\log \mathcal{N}(T, d, \epsilon)},$$

where d is the canonical metric $d(t, s) = (\mathbb{E}(X_t - X_s)^2)^{1/2}$.

A.2 One-dimensional case

A.2.1 Upper bound

We first derive an upper bound on the expected supremum of X_t^H introduced in Eq. (2).

Theorem 1 (Upper bound, 1-D case)**.** *Consider the stochastic process X_t^H defined by*

$$dX_t^H = \beta X_t dt + \sigma dB_t^H,$$

with $X_0 = 0, \sigma \in \mathbb{R}$ and $|\beta| < 1$. Then, for $t \in [0, 1]$,

$$\mathbb{E}\left[\sup_{t\in[0,1]} X_t^H\right] \leq C\left\{\sigma\sqrt{C_\beta} + \sigma\sqrt{\frac{C_\beta \pi}{H}}\right\},$$

where $C_\beta = e^{3\beta}$ and C are positive constants (independent of H).

Proof. We will use the bound on the supremum of a stochastic process X_t discussed in Theorem 6 that requires defining a metric δ on $[0, 1]$ such that

$$\mathbb{E}\left[\exp\left(\left(\frac{X_t^H - X_s^H}{\delta(t, s)}\right)^2\right)\right] \leq 2.$$

First, we note that $Z_{t,s} := X_t^H - X_s^H$ is known to be randomly distributed [3]. For X_0, we have $\mathbb{E}[Z_{t,s}] = 0$ since $\mathbb{E}[X_t^H] = X_0 e^{\beta t} = 0$. The variance of the increments of the process is derived in [3, 29] as

$$\begin{aligned}\mathrm{Var}[X_t^H - X_s^H] &= \mathrm{Var}[X_t^H] + \mathrm{Var}[X_s^H] - 2\mathrm{Cov}(X_t^H, X_s^H) \\ &= \sigma^2 H e^{-2\beta t}\int_0^{t-s} z^{2H-1}e^{\beta z}dz + \sigma^2 H e^{-2\beta s}\int_0^{t-s} z^{2H-1}e^{-\beta z}dz. \end{aligned} \tag{6}$$

Next, we upper bound the two integrals. We assume w.l.o.g. that β is positive. Else, we can simply exchange the bound on the two integrals (since we have $e^{-\beta z}$ in one integral and $e^{\beta z}$ in the other integral). Then we obtain

$$\int_0^{t-s} z^{2H-1}e^{\beta z}dz \leq e^{\beta}\int_0^{t-s} z^{2H-1}dz \tag{7}$$

$$\leq \frac{\bar{C}_\beta}{2H}(t-s)^{2H}, \tag{8}$$

where we used $e^{\beta z} \leq e^{\beta}$ in the first inequality since $(t-s) \leq 1$ (because $t \in [0,1]$) and $\bar{C}_\beta = e^{\beta}$ in the last inequality.

For the other integral,

$$\int_0^{t-s} z^{2H-1}e^{-\beta z}dz \leq \int_0^{t-s} z^{2H-1}dz \leq \frac{1}{2H}(t-s)^{2H}. \tag{9}$$

Therefore, for $s \leq t$,

$$\begin{aligned}\mathrm{Var}[X_t^H - X_s^H] &\leq \frac{1}{2}\sigma^2\bar{C}_\beta e^{-2\beta t}(t-s)^{2H} + \frac{1}{2}\sigma^2 e^{-2\beta s}(t-s)^{2H} \\ &\leq C_\beta\sigma^2(t-s)^{2H}, \end{aligned}\tag{10}$$

where we used $\bar{C}_\beta \geq 1$ and $C_\beta := e^{3\beta}$.

Let $Z \sim \mathcal{N}(0,1)$ and f_Z is the probability density function of Z. We have

$$\begin{aligned}\mathbb{E}\left[\exp\left(\frac{Z_{t,s}^2}{\delta^2(t,s)}\right)\right] &\leq \mathbb{E}\left[\exp\left(\frac{C_\beta\sigma^2(t-s)^{2H}Z^2}{\delta^2(t,s)}\right)\right] \\ &= \int_{-\infty}^{\infty} f_Z(z)\exp\left(\frac{C_\beta\sigma^2(t-s)^{2H}z^2}{\delta^2(t,s)}\right)dz \\ &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)\exp\left(\frac{C_\beta\sigma^2(t-s)^{2H}z^2}{\delta^2(t,s)}\right)dz \\ &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-z^2\left(\frac{1}{2} - \frac{C_\beta\sigma^2(t-s)^{2H}}{\delta^2(t,s)}\right)\right)dz. \end{aligned}$$

Recall that the integral of the Gaussian function $f(x) = e^{-ax^2}$ for $a > 0$ is $\int_{-\infty}^{\infty} e^{-ax^2}\,dx = \sqrt{\frac{\pi}{a}}$, therefore

$$\mathbb{E}\left[\exp\left(\frac{Z_{t,s}^2}{\delta^2(t,s)}\right)\right] = \frac{1}{\sqrt{2}}\left(\frac{1}{2} - \frac{C_\beta\sigma^2(t-s)^{2H}}{\delta^2(t,s)}\right)^{-\frac{1}{2}}, \tag{11}$$

where we need $\left(\frac{1}{2} - \frac{C_\beta\sigma^2(t-s)^{2H}}{\delta^2(t,s)}\right) > 0$.

Next, we choose δ such that

$$\begin{aligned} &\frac{1}{\sqrt{2}}\left(\frac{1}{2} - \frac{C_\beta\sigma^2(t-s)^{2H}}{\delta^2(t,s)}\right)^{-\frac{1}{2}} \leq 2 \\ \Longrightarrow &\left(\frac{1}{2} - \frac{C_\beta\sigma^2(t-s)^{2H}}{\delta^2(t,s)}\right) \geq \frac{1}{8} \\ \Longrightarrow &\frac{C_\beta\sigma^2(t-s)^{2H}}{\delta^2(t,s)} \leq \frac{3}{8} \end{aligned}$$

e.g.

$$\delta(t,s) = 2\sqrt{C_\beta}\sigma(t-s)^H.$$

Then

$$\begin{aligned}
U_\varepsilon^\delta(t_0) \quad &: \quad = \{t \in [0,1] : \delta(t,t_0) < \varepsilon\} \\
&= \quad \left\{ t \in [0,1] : |t-t_0| < \left(\frac{\varepsilon}{2\sigma\sqrt{C_\beta}}\right)^{\frac{1}{H}} \right\} \\
&= \quad U_r(t_0)
\end{aligned}$$

for $r := \left(\frac{\varepsilon}{2\sigma\sqrt{C_\beta}}\right)^{\frac{1}{H}}$. So

$$N_\delta(\varepsilon) \le \left[\left(\frac{\varepsilon}{2\sigma\sqrt{C_\beta}}\right)^{-\frac{1}{H}} \right] + 1 \text{ (} [\cdot] \text{ Gauss bracket).}$$

Hence

$$\begin{aligned}
\int_0^\infty \sqrt{\log(N_\delta(\varepsilon))}d\varepsilon = \int_0^{2\sigma\sqrt{C_\beta}} \sqrt{\log(N_\delta(\varepsilon))}d\varepsilon &\le \int_0^{2\sigma\sqrt{C_\beta}} \sqrt{\log\left(\left(\frac{\varepsilon}{2\sigma\sqrt{C_\beta}}\right)^{-\frac{1}{H}} + 1\right)}d\varepsilon \\
&\le \int_0^{2\sigma\sqrt{C_\beta}} \sqrt{\log\left(2\left(\frac{2\sigma\sqrt{C_\beta}}{\varepsilon}\right)^{\frac{1}{H}}\right)}d\varepsilon \\
&= \int_0^{2\sigma\sqrt{C_\beta}} \sqrt{\log(2) + \frac{1}{H}\log\left(\frac{2\sigma\sqrt{C_\beta}}{\varepsilon}\right)}d\varepsilon \\
&\le \int_0^{2\sigma\sqrt{C_\beta}} \sqrt{\log(2)} + \sqrt{\frac{1}{H}\log\left(\frac{2\sigma\sqrt{C_\beta}}{\varepsilon}\right)}d\varepsilon \\
&= 2\sqrt{\log(2)}\sigma\sqrt{C_\beta} + \sqrt{\frac{1}{H}}\int_0^{2\sigma\sqrt{C_\beta}} \sqrt{\log\left(\frac{2\sigma\sqrt{C_\beta}}{\varepsilon}\right)}d\varepsilon.
\end{aligned}$$

Then, using Theorem 6, we obtain that there exists a constant C_2 such that

$$\mathbb{E}\left[\sup_{t\in[0,1]} X_t^H\right] \le C_2 \left\{ 2\sqrt{\log(2)}\sigma\sqrt{C_\beta} + \sqrt{\frac{1}{H}}\int_0^{2\sigma\sqrt{C_\beta}} \sqrt{\log\left(\frac{2\sigma\sqrt{C_\beta}}{\varepsilon}\right)}d\varepsilon \right\}.$$

Finally, we obtain an analytical form for the integral using $\int_0^c \sqrt{\log(\frac{c}{x})}dx = c\frac{\sqrt{\pi}}{2}$ (which can be derived using $\int_0^1 \sqrt{\log(\frac{1}{x})}dx = \frac{\sqrt{\pi}}{2}$ combined with a change of variables).

□

A.2.2 Lower bound

We now turn our attention to a lower bound. This will demonstrate that the result obtained in the upper bound is actually sharp in terms of the dependency to the Hurst parameter H.

Theorem 2 (Lower bound, 1-D case)**.** *Consider the stochastic process X_t^H defined by*

$$dX_t^H = -\beta X_t dt + \sigma dB_t^H,$$

with $X_0 = 0, \sigma \in \mathbb{R}$ and $|\beta| < 1$. Then

$$\mathbb{E}\left[\sup_{t\in[0,1]} X_t^H\right] \geq c\left\{e^{-\frac{1}{2}}\sqrt{\frac{1}{H}\log\left(\frac{\sigma C_\beta}{e^{-\frac{1}{2}}}\right)}\right\},$$

where $C_\beta = e^{-2|\beta|}$ and c are positive constants (independent of H).

Proof. We will use Sudakov's inequality that relies on the canonical metric

$$\delta(t,s) = (\mathbb{E}(X_t^H - X_s^H)^2)^{1/2}.$$

In order to get an expression for $\delta(t,s)$, we first note that $Z_{t,s} = X_t^H - X_s^H$ is known to be randomly distributed [3]. For X_0, we have $\mathbb{E}[Z_{t,s}] = 0$ since $\mathbb{E}[X_t^H] = X_0 e^{\beta t} = 0$. The variance of the increments of the process is given in [3, 29] as

$$\begin{aligned}\text{Var}[X_t^H - X_s^H] &= \text{Var}[X_t^H] + \text{Var}[X_s^H] - 2\text{Cov}(X_t^H, X_s^H)\\ &= \sigma^2 H e^{-2\beta t}\int_0^{t-s} z^{2H-1}e^{\beta z}dz + \sigma^2 H e^{-2\beta s}\int_0^{t-s} z^{2H-1}e^{-\beta z}dz. \end{aligned} \tag{12}$$

Next, we lower bound the two integrals. We assume w.l.o.g. that β is positive. Else, we can simply exchange the bound on the two integrals (since we have $e^{-\beta z}$ in one integral and $e^{\beta z}$ in the other integral). Then we obtain

$$\begin{aligned}\int_0^{t-s} z^{2H-1}e^{\beta z}dz &= \int_0^{t-s} z^{2H-1}\left(1 + \beta z + \frac{\beta^2 z^2}{2} + \dots\right)dz\\ &= \frac{(t-s)^{2H}}{2H} + \beta\frac{(t-s)^{2H+1}}{2H+1} + \beta^2\frac{(t-s)^{2H+2}}{2H+2} + \dots\\ &\geq \frac{(t-s)^{2H}}{2H} + \beta\frac{(t-s)^{2H+1}}{2H+1}. \end{aligned} \tag{13}$$

For the other integral,

$$\int_0^{t-s} z^{2H-1}e^{-\beta z}dz \geq \int_0^{t-s} z^{2H-1}(1-\beta z)dz = \frac{(t-s)^{2H}}{2H} - \beta\frac{(t-s)^{2H+1}}{2H+1} \tag{14}$$

Therefore, for $s \leq t$,

$$\begin{aligned}\text{Var}[X_t^H - X_s^H] &\geq \sigma^2 H e^{-2\beta t}\left(\frac{(t-s)^{2H}}{2H} + \beta\frac{(t-s)^{2H+1}}{2H+1}\right)\\ &\quad + \sigma^2 H e^{-2\beta s}\left(\frac{(t-s)^{2H}}{2H} - \beta\frac{(t-s)^{2H+1}}{2H+1}\right)\\ &\geq \sigma^2 H e^{-2\beta t}\left(2\frac{(t-s)^{2H}}{2H}\right)\\ &= \sigma^2 e^{-2\beta t}(t-s)^{2H}\\ &= \sigma^2 C_\beta (t-s)^{2H}, \end{aligned} \tag{15}$$

where $C_\beta = e^{-2|\beta|}$.

We therefore obtain the following canonical metric

$$\delta(t,s) = (\mathbb{E}(X_t - X_s)^2)^{1/2} = \sigma C_\beta (t-s)^H.$$

Then

$$
\begin{aligned}
U_\varepsilon^\delta(t_0) \quad : \quad &= \{t \in [0,1] : \delta(t,t_0) < \varepsilon\} \\
&= \left\{ t \in [0,1] : |t - t_0| < \left(\frac{\varepsilon}{\sigma C_\beta} \right)^{\frac{1}{H}} \right\} \\
&= U_r(t_0),
\end{aligned}
$$

for $r := \left(\frac{\varepsilon}{\sigma C_\beta}\right)^{\frac{1}{H}}$. So

$$
N_\delta(\varepsilon) \geq \left[\left(\frac{\varepsilon}{\sigma C_\beta} \right)^{-\frac{1}{H}} \right].
$$

By Sudakov's inequality (Theorem 7), we have

$$
\mathbb{E}\left[\sup_{t \in [0,1]} X_t^H \right] \geq c \sup_{0 < \epsilon \leq \sigma C_\beta} \epsilon \sqrt{\log N_\delta(\varepsilon)} \tag{16}
$$

From our previous calculation, we have

$$
\begin{aligned}
\log N_\delta(\varepsilon) \geq \log \left(\frac{\sigma C_\beta}{\varepsilon} \right)^{\frac{1}{H}} &= \frac{1}{H} \log \left(\frac{\sigma C_\beta}{\varepsilon} \right) \\
&= \frac{1}{H} \big(\log(\sigma C_\beta) - \log(\varepsilon) \big)
\end{aligned}
$$

Since $\arg\max_{\varepsilon \in (0,1]} \varepsilon \sqrt{-\log(\varepsilon)} = e^{-\frac{1}{2}}$, we get

$$
\sup_{0 < \varepsilon \leq \sigma C_\beta} \varepsilon \sqrt{\log N_\delta(\varepsilon)} \geq e^{-\frac{1}{2}} \sqrt{\frac{1}{H} \log \left(\frac{\sigma C_\beta}{e^{-\frac{1}{2}}} \right)}.
$$

□

A.3 Multi-dimensional case

A.3.1 Lower bound

We first derive a lower bound on the expected supremum of X_t^H introduced in Eq. (2) in the multi-dimensional case.

Theorem 5 (Lower bound, d-dimensional case)**.** *Consider the d-dimensional stochastic process $X_t^H \in \mathbb{R}^d$ defined by*

$$
dX_t^H = AX_t dt + dB_t^H, X_0 = \xi,
$$

where $A \in \mathbb{R}^{d \times d}$ and $B_t^H = (B_t^{H,1}, \dots B_t^{H,d})$ is a fBM with $H \in (0, \frac{1}{2})$. Denote by $a_{ij}(t)$ the matrix entries of $\exp(tA)A$ and assume that there exists $i_0 \in \{1, \dots d\}$ such that $\min_{t \in [0,1]} a_{i_0 i_0}(t) \geq -1 + C_1$, where $0 < C_1 < 1$. Then

$$
\mathbb{E}\left[\sup_{t \in [0,1]} \|X_t^H - \mathbb{E}[X_t^H]\| \right] \geq c e^{-\frac{1}{2}} \sqrt{\frac{1}{H} \log \left(\frac{2C_1}{e^{-\frac{1}{2}}} \right)},
$$

where c is a constant independent of H.

Proof. The solution X_t (we omit the superscript H for simplicity) is given by

$$
X_t = \phi(t) \left(\xi + \int_0^t \phi^{-1}(s) A B_s^H ds \right) + B_t^H, \tag{17}
$$

where $\phi(t)$ is the fundamental solution, that solves the matrix equation

$$\dot{\phi}(t) = A\phi(t), \qquad \phi(0) = \text{Id}. \tag{18}$$

In the case where the matrix A is constant with time, the fundamental solution is given by

$$\phi(t) = e^{tA} = \sum_{k=0}^{\infty} \frac{(tA)^k}{k!}. \tag{19}$$

By centering X_t, we get

$$\tilde{X}_t := X_t - \mathbb{E}[X_t] = \int_0^t \underbrace{\phi(t)\phi^{-1}(s)}_{=\exp((t-s)A)} AB_s^H ds + B_t^H. \tag{20}$$

Using the stationarity of X_t, we find that for $t_1 \le t_2$,

$$\begin{aligned} \|\tilde{X}_{t_2} - \tilde{X}_{t_1}\|_{L^2(\Omega)} &= \|X_{t_2} - X_{t_1} - \mathbb{E}[X_{t_2} - X_{t_1}]\|_{L^2(\Omega)} \\ &= \|\tilde{X}_{t_2-t_1}\|_{L^2(\Omega)}. \end{aligned} \tag{21}$$

Denote by $a_{ij}(t)$ the matrix entries of $\exp(tA)A$. We have

$$\begin{aligned} \|\tilde{X}_{t_2-t_1}\|^2_{L^2(\Omega)} &= \mathbb{E}\left\| \int_0^{t_2-t_1} \exp((t_2-t_1-s)A)AB_s^H ds + B^H_{t_2-t_1} \right\|^2 \\ &= \sum_{i=1}^d \mathbb{E}\left[\left(\int_0^{t_2-t_1} \sum_{j=1}^d a_{ij}(t_2-t_1-s) B_s^{H,j} ds + B^{H,i}_{t_2-t_1} \right)^2 \right] \\ &= \sum_{i=1}^d \Bigg\{ \mathbb{E}\left[\left(\int_0^{t_2-t_1} \sum_{j=1}^d a_{ij}(t_2-t_1-s) B_s^{H,j} ds \right)^2 \right] + \mathbb{E}[(B^{H,i}_{t_2-t_1})^2] \\ &\qquad + 2\int_0^{t_2-t_1} \sum_{j=1}^d a_{ij}(t_2-t_1-s) \mathbb{E}[B_s^{H,j} B^{H,i}_{t_2-t_1}] ds \Bigg\} \\ &\overset{(i)}{=} \sum_{i=1}^d \Bigg\{ \underbrace{\mathbb{E}\left[\left(\int_0^{t_2-t_1} \sum_{j=1}^d a_{ij}(t_2-t_1-s) B_s^{H,j} ds \right)^2 \right]}_{=:A} + |t_2-t_1|^{2H} \\ &\qquad \underbrace{+ \int_0^{t_2-t_1} \sum_{j=1}^d a_{ij}(t_2-t_1-s) \left(|t_2-t_1|^{2H} + |s|^{2H} - |t_2-t_1-s|^{2H} \right) ds}_{=:B} \Bigg\} \\ &\ge \int_0^{t_2-t_1} a_{ii}(t_2-t_1-s) \left(|t_2-t_1|^{2H} + |s|^{2H} - |t_2-t_1-s|^{2H} \right) ds + |t_2-t_1|^{2H} \end{aligned}$$

where (i) uses $\mathbb{E}[(B^{H,i}_{t_2-t_1})^2] = |t_2 - t_1|^{2H}$ and $\mathbb{E}[B_s^{H,j} B^{H,i}_{t_2-t_1}] = \frac{1}{2}\left(|t_2-t_1|^{2H} + |s|^{2H} - |t_2-t_1-s|^{2H} \right)$.

Next, assume that there exists $i_0 \in \{1, \dots d\}$ such that

$$\min_{t \in [0,1]} a_{i_0 i_0}(t) \ge -1 + C_1,$$

where $0 \leq C_1 \leq 1$. Then

$$\begin{aligned}
\|\tilde{X}_{t_2} - \tilde{X}_{t_1}\|_{L^2(\Omega)} &\geq \|\tilde{X}^{i_0}_{t_2} - \tilde{X}^{i_0}_{t_1}\|_{L^2(\Omega)} \\
&\geq \int_0^{t_2-t_1} (-1+C_1)\left(|t_2-t_1|^{2H} + |s|^{2H} - |t_2-t_1-s|^{2H}\right) ds + |t_2-t_1|^{2H} \\
&= (-1+C_1)|t_2-t_1|^{2H+1} + |t_2-t_1|^{2H} \\
&\geq (-1+C_1)|t_2-t_1|^{2H} + |t_2-t_1|^{2H} \\
&\geq C_1|t_2-t_1|^{2H}.
\end{aligned}$$

Taking $\delta(t_2,t_1) = C_1|t_2-t_1|^H$ as the metric on $[0,1]$, we obtain

$$\begin{aligned}
U^\delta_\varepsilon(t_0) \quad : &\quad = \{t \in [0,1] : \delta(t,t_0) < \varepsilon\} \\
&= \quad \left\{ t \in [0,1] : |t-t_0| < \left(\frac{\varepsilon}{C_1}\right)^{\frac{1}{H}} \right\} \\
&= \quad U_r(t_0)
\end{aligned}$$

for $r := \left(\frac{\varepsilon}{C_1}\right)^{\frac{1}{H}}$. So

$$N_\delta(\varepsilon) \geq \left[\left(\frac{\varepsilon}{C_1}\right)^{-\frac{1}{H}} \right].$$

By Sudakov's inequality (Theorem 7), we have

$$\mathbb{E}\left[\sup_{t\in[0,1]} (\pm \tilde{X}^{i_0}_t) \right] \geq c \sup_{0<\epsilon\leq 2C_1} \epsilon\sqrt{\log N_\delta(\varepsilon)}.$$

where C_2 is a constant independent of H.

From our previous calculation, we have

$$\begin{aligned}
\log N_\delta(\varepsilon) \geq \log\left(\frac{2C_1}{\varepsilon}\right)^{\frac{1}{H}} &= \frac{1}{H}\log\left(\frac{2C_1}{\varepsilon}\right) \\
&= \frac{1}{H}\big(\log(2C_1) - \log(\varepsilon)\big).
\end{aligned}$$

Since $\arg\max_{\varepsilon\in(0,1]} \varepsilon\sqrt{-\log(\varepsilon)} = e^{-\frac{1}{2}}$, we get

$$\sup_{0<\varepsilon\leq 2C_1} \varepsilon\sqrt{\log N_\delta(\varepsilon)} \geq e^{-\frac{1}{2}} \sqrt{\frac{1}{H}\log\left(\frac{2C_1}{e^{-\frac{1}{2}}}\right)}.$$

We conclude that

$$\mathbb{E}\left[\sup_{t\in[0,1]} \|\tilde{X}_t\| \right] \geq c e^{-\frac{1}{2}} \sqrt{\frac{1}{H}\log\left(\frac{2C_1}{e^{-\frac{1}{2}}}\right)}.$$

□

Alternative proof for lower bound Recall we have shown that

$$\begin{aligned}
\|\tilde{X}_{t_2-t_1}\|^2_{L^2(\Omega)} = \sum_{i=1}^d \Bigg\{ &\underbrace{\mathbb{E}\left[\left(\int_0^{t_2-t_1} \sum_{j=1}^d a_{ij}(t_2-t_1-s) B^{H,j}_s ds \right)^2 \right]}_{=:A} + |t_2-t_1|^{2H} \\
&+ \underbrace{\int_0^{t_2-t_1} \sum_{j=1}^d a_{ij}(t_2-t_1-s)\left(|t_2-t_1|^{2H} + |s|^{2H} - |t_2-t_1-s|^{2H}\right) ds}_{=:B} \Bigg\}.
\end{aligned}$$

Let's first focus on the term A:

$$
\begin{aligned}
&\mathbb{E}\left[\left(\int_0^{t_2-t_1}\sum_{j=1}^d a_{ij}(t_2-t_1-s)B_s^{H,j}ds\right)^2\right]\\
&\quad=\mathbb{E}\left[\left(\int_0^{t_2-t_1}\sum_{j=1}^d a_{ij}(t_2-t_1-s_1)B_{s_1}^{H,j}ds_1\right)\left(\int_0^{t_2-t_1}\sum_{j=1}^d a_{ij}(t_2-t_1-s_2)B_{s_2}^{H,j}ds_2\right)\right]\\
&\quad=\int_0^{t_2-t_1}\int_0^{t_2-t_1}\mathbb{E}\left[\left(\sum_{j=1}^d a_{ij}(t_2-t_1-s_1)B_{s_1}^{H,j}\right)\left(\sum_{j=1}^d a_{ij}(t_2-t_1-s_2)B_{s_2}^{H,j}\right)\right]ds_1ds_2\\
&\quad=2\int_0^{t_2-t_1}\int_0^{s_2}\mathbb{E}\left[\left(\sum_{j=1}^d a_{ij}(t_2-t_1-s_1)B_{s_1}^{H,j}\right)\left(\sum_{j=1}^d a_{ij}(t_2-t_1-s_2)B_{s_2}^{H,j}\right)\right]ds_1ds_2\\
&\quad=2\int_0^{t_2-t_1}\int_0^{s_2}\sum_{j=1}^d a_{ij}(t_2-t_1-s_1)a_{ij}(t_2-t_1-s_2)\mathbb{E}\left[B_{s_1}^{H,j}B_{s_2}^{H,j}\right]ds_1ds_2\\
&\quad=\int_0^{t_2-t_1}\int_0^{s_2}\sum_{j=1}^d a_{ij}(t_2-t_1-s_1)a_{ij}(t_2-t_1-s_2)(|s_2|^{2H}+|s_1|^{2H}+|s_2-s_1|^{2H})ds_1ds_2\\
&\quad\geq d\min_{j=1}^d\min_{u_1,u_2\in[0,1]}a_{ij}(u_1)a_{ij}(u_2)\int_0^{t_2-t_1}\int_0^{s_2}(|s_2|^{2H}+|s_1|^{2H}+|s_2-s_1|^{2H})ds_1ds_2.
\end{aligned}
$$

Then, note that

$$
\int_0^{s_2}(|s_2|^{2H}+|s_1|^{2H}+|s_2-s_1|^{2H})ds_1=|s_2|^{2H+1}+\frac{|s_2|^{2H+1}}{2H+1}-\frac{|s_2|^{2H+1}}{2H+1}=|s_2|^{2H+1}.
$$

Therefore

$$
\begin{aligned}
\mathbb{E}\left[\left(\int_0^{t_2-t_1}\sum_{j=1}^d a_{ij}(t_2-t_1-s)B_s^{H,j}ds\right)^2\right]&\geq d\min_{j=1}^d\min_{u_1,u_2\in[0,1]}a_{ij}(u_1)a_{ij}(u_2)\int_0^{t_2-t_1}|s_2|^{2H+1}ds_2\\
&=d\min_{j=1}^d\min_{u_1,u_2\in[0,1]}a_{ij}(u_1)a_{ij}(u_2)\frac{|t_2-t_1|^{2H+2}}{2H+2}.
\end{aligned}
$$

For the term B, we have

$$
\begin{aligned}
&\int_0^{t_2-t_1}\sum_{j=1}^d a_{ij}(t_2-t_1-s)\left(|t_2-t_1|^{2H}+|s|^{2H}-|t_2-t_1-s|^{2H}\right)ds\\
&\qquad\geq d\min_{j=1}^d\min_{u\in[0,1]}a_{ij}(u)\int_0^{t_2-t_1}\left(|t_2-t_1|^{2H}+|s|^{2H}-|t_2-t_1-s|^{2H}\right)ds\\
&\qquad\geq d\min_{j=1}^d\min_{u\in[0,1]}a_{ij}(u)|t_2-t_1|^{2H+1}.
\end{aligned}
$$

Combining the lower bounds on A and B, we obtain

$$
\begin{aligned}
&\|\tilde{X}_{t_2-t_1}\|^2_{L^2(\Omega)}\\
&\quad\geq d\min_{j=1}^d\min_{u_1,u_2\in[0,1]}a_{ij}(u_1)a_{ij}(u_2)\frac{|t_2-t_1|^{2H+2}}{2H+2}+d\min_{j=1}^d\min_{u\in[0,1]}a_{ij}(u)|t_2-t_1|^{2H+1}+|t_2-t_1|^{2H}\\
&\quad=\left(\underbrace{\frac{d}{2H+2}\min_{j=1}^d\min_{u_1,u_2\in[0,1]}a_{ij}(u_1)a_{ij}(u_2)}_{=:a_1(i)}|t_2-t_1|^2+\underbrace{d\min_{j=1}^d\min_{u\in[0,1]}a_{ij}(u)}_{=:a_2(i)}+1\right)|t_2-t_1|^{2H}.
\end{aligned}
$$

Let $C_1 = \min_{t\in[0,1]}(a_1(i)t^2 + a_2(i)t + 1) > 0$. We can then choose the metric on $[0, 1]$ as $\delta(t_2, t_1) = C_1|t_2 - t_1|^H$.

The rest of the proof is identical but we obtain a different constant C_1 that covers different scenarios of the entries of the matrix A.

A.3.2 Upper bound

We now turn our attention to an upper bound. This will demonstrate that the result obtained in the lower bound is actually sharp in terms of the dependency to the Hurst parameter H.

Theorem 4 (Upper bound, d-dimensional case)**.** *Consider the d-dimensional stochastic process $X_t^H \in \mathbb{R}^d$ defined by*

$$dX_t^H = AX_t dt + dB_t^H, X_0 = \xi,$$

where $A \in \mathbb{R}^{d\times d}$ and $B_t^H = (B_t^{H,1}, \dots B_t^{H,d})$ is a fBM with $H \in (0, 1)$. Denote by $a_{ij}(t)$ the matrix entries of $\exp(tA)A$. Then

$$\mathbb{E}\left[\sup_{t\in[0,1]} \|X_t^H - \mathbb{E}[X_t^H]\|\right] \leq Cd^2C_2 + \frac{Cd^2C_2\sqrt{\pi}}{2\sqrt{H}},$$

where $C_2 = \max_{i,j=1}^d \sup_{t\in[0,1]} |a_{ij}(t)|^2 + \max_{i,j=1}^d \sup_{t\in[0,1]} |a_{ij}(t)| + 1$ and $C > 0$ is a constant independent of H.

Proof. The solution X_t (we omit the superscript H for simplicity) is given by

$$X_t = \phi(t)\left(\xi + \int_0^t \phi^{-1}(s)AB_s^H ds\right) + B_t^H, \tag{22}$$

where $\phi(t)$ is the fundamental solution, that solves the matrix equation

$$\dot{\phi}(t) = A\phi(t), \qquad \phi(0) = \text{Id}. \tag{23}$$

In the case where the matrix A is constant with time, the fundamental solution is given by

$$\phi(t) = e^{tA} = \sum_{k=0}^{\infty} \frac{(tA)^k}{k!}. \tag{24}$$

By centering X_t, we get

$$\tilde{X}_t := X_t - \mathbb{E}[X_t] = \int_0^t \underbrace{\phi(t)\phi^{-1}(s)}_{=\exp((t-s)A)} AB_s^H ds + B_t^H. \tag{25}$$

Using the stationarity of X_t, we find that for $t_1 \leq t_2$,

$$\begin{aligned} \|\tilde{X}_{t_2} - \tilde{X}_{t_1}\|_{L^2(\Omega)} &= \|X_{t_2} - X_{t_1} - \mathbb{E}[X_{t_2} - X_{t_1}]\|_{L^2(\Omega)} \\ &= \|\tilde{X}_{t_2-t_1}\|_{L^2(\Omega)}. \end{aligned} \tag{26}$$

Denote by $a_{ij}(t)$ the matrix entries of $\exp(tA)A$. We have

$$
\begin{aligned}
\|\tilde{X}_{t_2-t_1}\|^2_{L^2(\Omega)} &= \mathbb{E}\left\|\int_0^{t_2-t_1}\exp((t_2-t_1-s)A)AB^H_s\,ds + B^H_{t_2-t_1}\right\|^2 \\
&= \sum_{i=1}^d \mathbb{E}\left[\left(\int_0^{t_2-t_1}\sum_{j=1}^d a_{ij}(t_2-t_1-s)B^{H,j}_s\,ds + B^{H,i}_{t_2-t_1}\right)^2\right] \\
&= \sum_{i=1}^d \Big\{\mathbb{E}\left[\left(\int_0^{t_2-t_1}\sum_{j=1}^d a_{ij}(t_2-t_1-s)B^{H,j}_s\,ds\right)^2\right] + \mathbb{E}[(B^{H,i}_{t_2-t_1})^2] \\
&\quad + 2\int_0^{t_2-t_1}\sum_{j=1}^d a_{ij}(t_2-t_1-s)\mathbb{E}[B^{H,j}_s B^{H,i}_{t_2-t_1}]\,ds\Big\} \\
&\overset{(i)}{=} \sum_{i=1}^d \Big\{\underbrace{\mathbb{E}\left[\left(\int_0^{t_2-t_1}\sum_{j=1}^d a_{ij}(t_2-t_1-s)B^{H,j}_s\,ds\right)^2\right]}_{=:A} + |t_2-t_1|^{2H} \\
&\quad + \underbrace{\int_0^{t_2-t_1}\sum_{j=1}^d a_{ij}(t_2-t_1-s)\left(|t_2-t_1|^{2H}+|s|^{2H}-|t_2-t_1-s|^{2H}\right)ds}_{=:B}\Big\},
\end{aligned}
$$

where (i) uses $\mathbb{E}[(B^{H,i}_{t_2-t_1})^2] = |t_2 - t_1|^{2H}$ and $\mathbb{E}[B^{H,j}_s B^{H,i}_{t_2-t_1}] = \frac{1}{2}\left(|t_2-t_1|^{2H}+|s|^{2H}-|t_2-t_1-s|^{2H}\right)$.

We will once again build on the proof in the 1-dimensional case. To do so, we first observe that, given an interval $T = [0,1]$,

$$
\begin{aligned}
\sup_{t\in T}\|\tilde{X}_t\| &\le \sup_{t\in T}|\tilde{X}^1_t| + \cdots + \sup_{t\in T}|\tilde{X}^d_t| \\
&\le \sup_{t\in T}\tilde{X}^1_t + \sup_{t\in T}-\tilde{X}^1_t + \cdots + \sup_{t\in T}\tilde{X}^d_t + \sup_{t\in T}-\tilde{X}^d_t,
\end{aligned}
$$

thus

$$
\mathbb{E}\left[\sup_{t\in T}\|\tilde{X}_t\|\right] \le \mathbb{E}\left[\sup_{t\in T}\tilde{X}^1_t\right] + \mathbb{E}\left[\sup_{t\in T}-\tilde{X}^1_t\right] + \cdots + \mathbb{E}\left[\sup_{t\in T}\tilde{X}^d_t\right] + \mathbb{E}\left[\sup_{t\in T}-\tilde{X}^d_t\right].
$$

Focusing on the i-th coordinate, we first get for the term A that

$$
\begin{aligned}
&\mathbb{E}\left[\left(\int_0^{t_2-t_1}\sum_{j=1}^d a_{ij}(t_2-t_1-s)B_s^{H,j}ds\right)^2\right]\\
&=\mathbb{E}\left[\left(\int_0^{t_2-t_1}\sum_{j=1}^d a_{ij}(t_2-t_1-s_1)B_{s_1}^{H,j}ds_1\right)\left(\int_0^{t_2-t_1}\sum_{j=1}^d a_{ij}(t_2-t_1-s_2)B_{s_2}^{H,j}ds_2\right)\right]\\
&=\int_0^{t_2-t_1}\int_0^{t_2-t_1}\mathbb{E}\left[\left(\sum_{j=1}^d a_{ij}(t_2-t_1-s_1)B_{s_1}^{H,j}\right)\left(\sum_{j=1}^d a_{ij}(t_2-t_1-s_2)B_{s_2}^{H,j}\right)\right]ds_1ds_2\\
&=2\int_0^{t_2-t_1}\int_0^{s_2}\mathbb{E}\left[\left(\sum_{j=1}^d a_{ij}(t_2-t_1-s_1)B_{s_1}^{H,j}\right)\left(\sum_{j=1}^d a_{ij}(t_2-t_1-s_2)B_{s_2}^{H,j}\right)\right]ds_1ds_2\\
&=2\int_0^{t_2-t_1}\int_0^{s_2}\sum_{j=1}^d a_{ij}(t_2-t_1-s_1)a_{ij}(t_2-t_1-s_2)\mathbb{E}\left[B_{s_1}^{H,j}B_{s_2}^{H,j}\right]ds_1ds_2\\
&=\int_0^{t_2-t_1}\int_0^{s_2}\sum_{j=1}^d a_{ij}(t_2-t_1-s_1)a_{ij}(t_2-t_1-s_2)(|s_2|^{2H}+|s_1|^{2H}+|s_2-s_1|^{2H})ds_1ds_2\\
&\leq d\max_{j=1}^d\sup_{t\in[0,1]}|a_{ij}(t)|^2\int_0^{t_2-t_1}\int_0^{s_2}(|s_2|^{2H}+|s_1|^{2H}+|s_2-s_1|^{2H})ds_1ds_2.
\end{aligned}
$$

Then, note that

$$
\int_0^{s_2}(|s_2|^{2H}+|s_1|^{2H}+|s_2-s_1|^{2H})ds_1=|s_2|^{2H+1}+\frac{|s_2|^{2H+1}}{2H+1}-\frac{|s_2|^{2H+1}}{2H+1}=|s_2|^{2H+1}.
$$

Therefore

$$
\begin{aligned}
\mathbb{E}\left[\left(\int_0^{t_2-t_1}\sum_{j=1}^d a_{ij}(t_2-t_1-s)B_s^{H,j}ds\right)^2\right]&\leq d\max_{j=1}^d\sup_{t\in[0,1]}|a_{ij}(t)|^2\int_0^{t_2-t_1}|s_2|^{2H+1}ds_2\\
&=d\max_{j=1}^d\sup_{t\in[0,1]}|a_{ij}(t)|^2\frac{|t_2-t_1|^{2H+2}}{2H+2}.
\end{aligned}
$$

For the term B, we have

$$
\int_0^{t_2-t_1}\sum_{j=1}^d a_{ij}(t_2-t_1-s)\left(|t_2-t_1|^{2H}+|s|^{2H}-|t_2-t_1-s|^{2H}\right)ds\leq d\max_{j=1}^d\sup_{t\in[0,1]}|a_{ij}(t)||t_2-t_1|^{2H+1}
$$

Combining the bound on the terms A and B, we obtain the following

$$
\begin{aligned}
\|\tilde{X}_{t_2}^i-\tilde{X}_{t_1}^i\|_{L^2(\Omega)}^2&\leq d\max_{i,j=1}^d\sup_{t\in[0,1]}|a_{ij}(t)|^2\frac{|t_2-t_1|^{2H+2}}{2H+2}+d\max_{i,j=1}^d\sup_{t\in[0,1]}|a_{ij}(t)||t_2-t_1|^{2H+1}+|t_2-t_1|^{2H}\\
&\leq dC_2|t_2-t_1|^{2H},
\end{aligned}
$$

where $C_2=\max_{i,j=1}^d\sup_{t\in[0,1]}|a_{ij}(t)|^2+\max_{i,j=1}^d\sup_{t\in[0,1]}|a_{ij(t)}|+1$.

Taking $\delta^2(t_2,t_1)=dC_2|t_2-t_1|^{2H}$ as the metric on $[0,1]$, we obtain

$$
\begin{aligned}
U_\varepsilon^\delta(t_0)\quad:&=\{t\in[0,1]:\delta(t,t_0)<\varepsilon\}\\
&=\left\{t\in[0,1]:|t-t_0|<\left(\frac{\varepsilon}{dC_2}\right)^{\frac{1}{H}}\right\}\\
&=U_r(t_0)
\end{aligned}
$$

for $r := \left(\frac{\varepsilon}{dC_2}\right)^{\frac{1}{H}}$. So

$$N_\delta(\varepsilon) \leq \left[\left(\frac{\varepsilon}{dC_2}\right)^{-\frac{1}{H}}\right] + 1.$$

Hence

$$\begin{aligned}
\int_0^\infty \sqrt{\log(N_\delta(\varepsilon))} d\varepsilon = \int_0^{dC_2} \sqrt{\log(N_\delta(\varepsilon))} d\varepsilon \leq \int_0^{dC_2} \sqrt{\log\left(\left(\frac{\varepsilon}{dC_2}\right)^{-\frac{1}{H}} + 1\right)} d\varepsilon \\
&\leq \int_0^{dC_2} \sqrt{\log\left(2\left(\frac{dC_2}{\varepsilon}\right)^{\frac{1}{H}}\right)} d\varepsilon \\
&= \int_0^{dC_2} \sqrt{\log(2) + \frac{1}{H}\log\left(\frac{dC_2}{\varepsilon}\right)} d\varepsilon \\
&\leq \int_0^{dC_2} \sqrt{\log(2)} + \sqrt{\frac{1}{H}\log\left(\frac{dC_2}{\varepsilon}\right)} d\varepsilon \\
&= dC_2\sqrt{\log(2)} + \sqrt{\frac{1}{H}} \int_0^{dC_2} \sqrt{\log\left(\frac{dC_2}{\varepsilon}\right)} d\varepsilon.
\end{aligned}$$

Then, using Theorem 6 applied to each centered component process (with positive and negative sign), we obtain that there exists a constant C such that

$$\mathbb{E}\left[\sup_{t\in T} \|\tilde{X}_t\|\right] \leq C2d^2C_2\sqrt{\log(2)} + C2d\sqrt{\frac{1}{H}} \int_0^{dC_2} \sqrt{\log\left(\frac{dC_2}{\varepsilon}\right)} d\varepsilon.$$

Finally, we obtain an analytical form for the integral using $\int_0^c \sqrt{\log(\frac{c}{x})} dx = c\frac{\sqrt{\pi}}{2}$ (which can be derived using $\int_0^1 \sqrt{\log(\frac{1}{x})} dx = \frac{\sqrt{\pi}}{2}$ combined with a change of variables).

□

B Additional experiments

B.1 Demonstration of the $1/\sqrt{H}$ scaling from Theorems 1 and 2

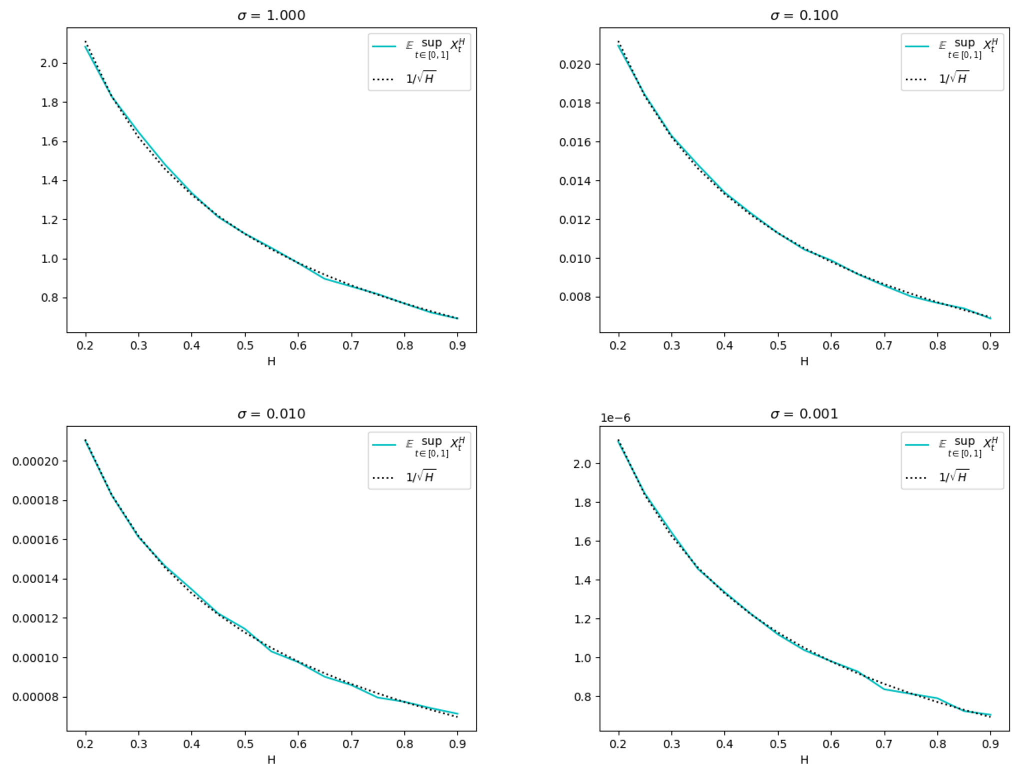

Figure 5: Demonstration of the $1/\sqrt{H}$ scaling of the expectation of the fOUP, as shown in Thms. 1 and 2. Drift parameter β is set to $\beta = 1/2$. Each plot for a different value of $\sigma \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}\}$. For all values of σ, we get an almost perfect fit. More details in text.

In this section, we demonstrate the $1/\sqrt{H}$ scaling of the expected supremum of the one-dimensional fOUP. Recall that Theorem 1 proved that

$$\mathbb{E}\left[\sup_{t\in[0,1]} \|X_t^H\|\right] \leq C\frac{1}{\sqrt{H}},$$

for $C > 0$. Theorem 2 and Corollary 3 also provide a lower bound

$$\mathbb{E}\left[\sup_{t\in[0,1]} \|X_t^H\|\right] \geq C\frac{1}{\sqrt{H}},$$

although they either require 1) $\sigma > e^{2|\beta|\frac{1}{2}}$ and $|\beta| < 1$ or 2) $0 < \sigma < e^{2|\beta|-\frac{1}{2}}$ and $|\beta| \leq \frac{1}{4}$.

Here, we want to verify whether the lower bound $\mathbb{E}\left[\sup_{t\in[0,1]} \|X_t^H\|\right] \sim \frac{1}{\sqrt{H}}$ holds generally, even outside the range required for σ and β. To this end, we ran a Monte Carlo simulation with 10^3 simulations of this expectation and plotted it next to a function of the form $f(x) = w_0 + w_1\frac{1}{\sqrt{H}}$, where we fitted (w_0, w_1) by ordinary least squares. For the discretization, we used 10^5 steps (i.e. a step size η of 10^{-5}). For H, we chose an evenly spaced grid of mesh size 0.05 from 0.2 to 0.9.[3]

[3] We did not include $H = 0.1$ here because the Monte-Carlo approximation is known not be a good approximation for such small values of H. This issue is described in more detail in [8] (see e.g. Figure 1 and description in the text). We remark that we still observe a relatively good fit (slightly worse than $H = 0.2$) despite the poorer quality of the approximation; the relative errors in the four settings of Fig. B.1 are $(0.19, 0.10, 0.17, 0.17)$ then.

The resulting plots are depicted in Fig. B.1 for noise scales $\sigma \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}\}$. In each plot, we used $\beta = 1/2$ such that all values of σ violate both conditions: 1) $\sigma > e^{2|\beta|-1/2}$ and $|\beta| < 1$ or 2) $0 < \sigma < e^{2|\beta|-\frac{1}{2}}$ and $|\beta| \leq \frac{1}{4}$. . Nonetheless, the fit between the simulation and the theoretical scaling is almost perfect. This confirms our theoretical findings and shows that the scaling also holds outside the regime of Theorem 2 and Corollary 3.

B.2 Justification for choice of σ in Section 5.2

In this section, we justify the use of $\sigma = 0.005$ in all experiments from Section 5.2. There, we used $\sigma = 0.005$ for all $H \in \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5\}$ in Figure 4, i.e. in both the Markovian $\{0.0, 0.5\}$ and the non-Markovian cases $\{0.1, 0.2, 0.3, 0.4\}$. Here, we explain why σ is not chosen dependent on H.

Our justification is simple: The choice of $\sigma = 0.005$ is a good choice for all H. To see this, consider Figure 6. For a fixed $H \in \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5\}$ and all $\sigma \in \{0.05, 0.005., 0.0005, 0.00005\}$, they depict the performance on the regularized quadratic from Section 5.2, analogous to Figure 4.

We observe that indeed, for all choices of H, the curve for $\sigma = 0.005$ is best – in the sense that it exits the saddle fastest (by reaching a negative value) and converges quickly to the final minimum. Clearly, for all H, there is no point in picking an even smaller or larger order of magnitude for σ. Hence, we conclude that the choice of $\sigma = 0.005$ is close to the optimal choice for all H. While we acknowledge that grid-searching the optimal σ at a finer scale would maybe uncover small differences in the optimal σ, such precision would be beyond our goal of a proof-of-concept and would distract from the bigger picture. In fact, the constant σ also ensures that the increments of fBM with different H are distributed the same (but correlated differently). Thus, this experimental set-up allows us to zoom in on the effect of the Hurst parameter H – with the result that the non-Markovian $H = 0.3$ performs best.

B.3 Moving to better local minima in bi-stable potentials

We consider the behavior of fPGD in a bi-stable optimization landscape, given by

$$f(x) = \left[v_0 + \frac{k_0}{2}x^2\right]\chi_{x\leq a} + \left[v_1 - \frac{k_1}{2}\left(x - \frac{a+c}{2}\right)^2\right]\chi_{a<x\leq c} + \left[v_2 + \frac{k_2}{2}(x-m)^2\right]\chi_{x>c}, \tag{27}$$

where χ is the indicator function and the parameters (v_0, v_1, v_2, a, c, m) determine the position and shape of two adjacent local minima. Similar landscapes have been studied in the context of escaping spurious local minima [55] and in the context of generalization in machine learning by selecting flat minima [33, 53], which are by some believed to generalized better. We here study both cases. For the former case, we set $(v_0, v_1, v_2, a, c, m) = (0.0, 45.0, -12.5, 3.5, 13.9, 18.0)$; for the corresponding landscape see the upper left subplot of Fig. 7. For the latter case, we set $(v_0, v_1, v_2, a, c, m) = (0.0, 30.0, 0.0, 1.2, 8.7, 15.0)$; for the for the corresponding landscape see the upper left subplot of Fig. 8.

On these landscapes, we run fPGD with different choices of H. Note that, in the two minima (and the saddle in between at $(a+c)/2$) fPGD is indeed a discretization of a fractional Ornstein–Uhlenbeck process, like the one considered in Theorems 1 and 2. We thus expect that a small Hurst parameter leads to a faster escape from the shallow minimum (left) to the deep minimum (right) in Fig. 7, or from the sharp to the flat minimum in Fig. 8. In both cases, we run fPGD for $N = 1000$ iterations with a fixed step size of $\eta = 1.0$, initialised at $x = 0$ (the bottom of the shallow/sharp minimum). To stay in the setting of Theorems 1 and 2, we set $\sigma = 80.0 \cdot N^{-H}$. Thus, for all $H \in (0, 1)$, the employed fBM B_t^H, defined on the time interval $t \in [0, N]$, is by Rmk. 1 distributed like $80.0 \cdot B_t^H$ on $t \in [0, 1]$ – in the sense that the discrete steps $\{B_k^H; k = 0, \ldots, N\}$ are distributed like $\{80.0 \cdot B_{k/N}^H; k = 0, \ldots, N\}$. By this H-dependent choice of σ, we give all fBMs the same final variance of 80.0^2 which lets us zoom in on the impact of correlation. (Otherwise the perturbations would have different variances; this way we disentangle the correlation from the variance.) We note that here the experimental outcomes are not sensitive to the precise value of σ.

Figs. 7 and 8 depict our findings. We say that fPGD is in the shallow/sharp minimum at iteration k if its iterate x_k is smaller than a, and in the deep/flat minimum if it is larger than c; both a and c are plotted as dashed lines in the optimization landscape (top row) and in the trajectories (bottom row).

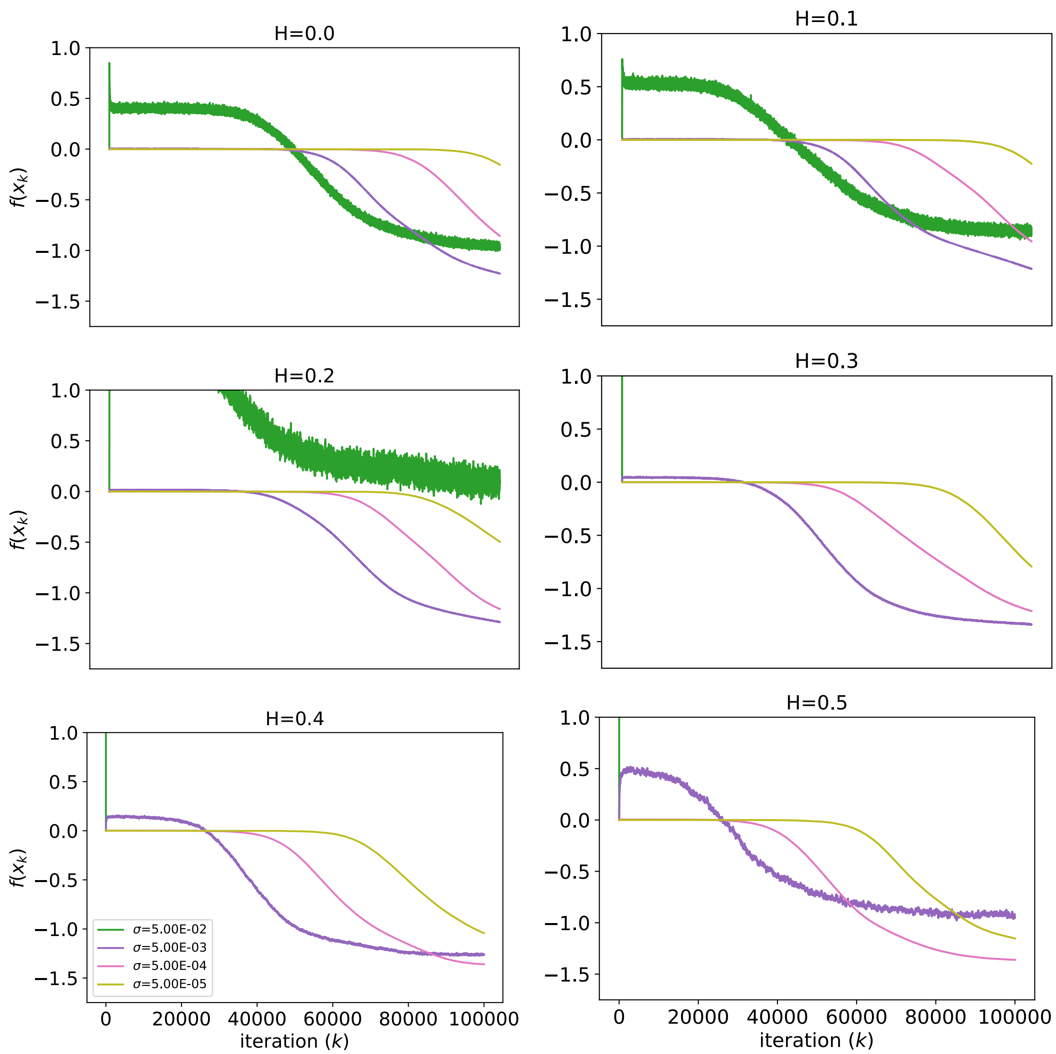

Figure 6: Loss curves on the embedded saddle landscape, analogous to Figure 4. Each subplot shows the loss curves for a fixed value of H and different values of σ. For all H, the purple curve is the best, which is the $\sigma = 0.005$ used in the main paper. (For the case $H = 0.5$, maybe the pink curve is equally good, but not better than the curves for $H = 0.3$ or $H = 0.4$.) Our claim that the non-Markovian $H \in \{0.3, 0.4\}$ perform best in this setting is therefore confirmed.

In the bottom row, for all H, we start below the lower dashed line a (shallow/sharp minimum) and then cross above the higher dashed line c (deep/flat minimum). It is evident that a smaller H enables faster crossings. We thus escape faster from the spurious shallow/sharp minimum to the desirable deep/flat one. (Note, however, if the noise is not cooled down, one might oscillate back.) The upper right plot depicts the cumulative distribution function of the first-exit time (i.e. the first time fPGD reaches the deep/flat minimum, above c). Again, we can observe that for a small Hurst parameter the probability of a fast exit is significantly higher.

Overall, these findings for fPGD match the continuous-time result of Theorems 1 and 2. Note that physicists [45] have also observed faster escapes for smaller Hurst parameters on a related Kramer-like escape problem.

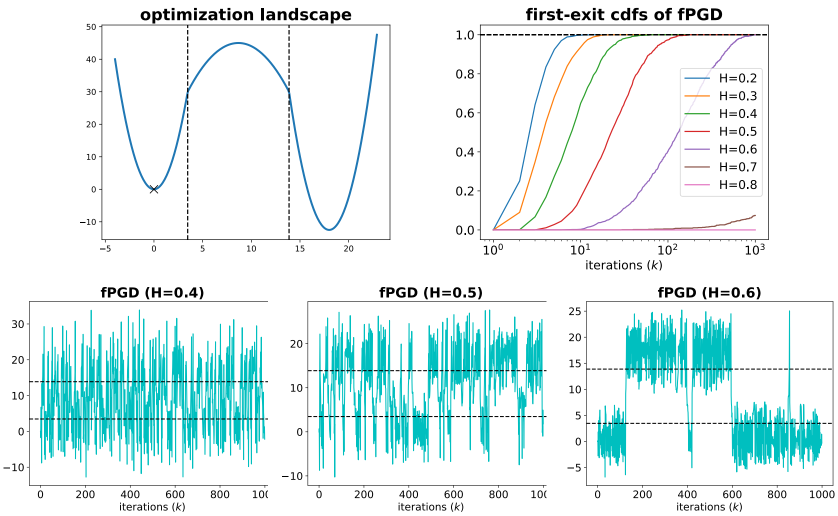

Figure 7: Experiments with fPGD (4) on the bi-stable objective (27) initialised at shallow minimum $x = 0$. Objective depicted on the upper right. Behavior of fPGD for $H \in \{0.4, 0.5, 0.6\}$ in bottom row. Smaller H gives faster switching between minima. Faster first-exit times from the shallow to the deep minimum are demonstrated by the cumulative density functions (cdf), depicted on the upper right. The cdf is computed by averaging over 1000 simulations. Details in text.

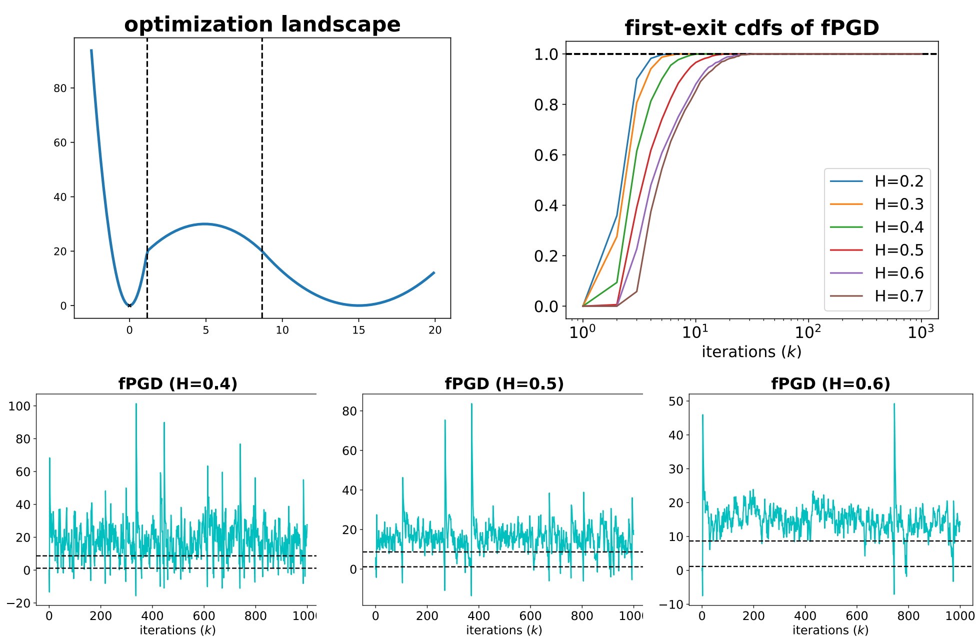

Figure 8: Figure analogous to Figure 7, but with two minima of equal depth but different width. Again, we find that a small Hurst parameters lead to faster fluctuations between the two minima (bottom row). Also, a small H gives faster first-exit times from the sharp to the flat minimum (upper left plot). All algorithmic parameters are as in Figure 7.